

漸進的な話し言葉翻訳における翻訳処理単

渡邊 善之

松原 茂樹

外山 勝彦

稲垣 康善 (名古屋)

1 はじめに

著者らが提案した漸進的な英日話し言葉翻訳手法は、英語入力に対してできる限り語単位で解析・変換処理を実行する [1]。このため、日本語翻訳結果の即時的な生成が可能であり、入力と出力との間の高度同時進行性を実現している。しかし、翻訳処理における即時性の度合と翻訳精度との間には、一般にトレードオフが存在する。より品質の高い翻訳処理を実現するために、語単位からそれ以上に処理単位を緩めることは一つの方法である。本稿では、漸進的な話し言葉翻訳における効果的な翻訳処理単位について実験的考察を与える。

2 言い直しの生成とその削減手法

漸進的な英日話し言葉翻訳システムは、語が入力されるたびにそれまでの入力に対する構文構造を作成し(解析部)、それに対して変換規則を適用する(変換部)ことにより日本語翻訳結果を作り上げる [2]。構文的曖昧性により複数の構文構造が作成された場合にも、翻訳処理の即時性を維持するために敢えて一つの構造を選択し、それに対してのみ変換処理を実行する。構文構造の選択を誤れば出力される翻訳結果も誤ることになるが、それが判明した時点で正しい構文構造を再度選択すれば、正しく言い直すことが可能となる。例えば、文

(2.1) Ken met her aunt in the park yesterday.

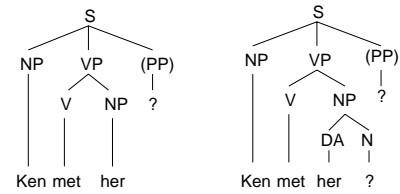
の“her”が入力された時点では、その構文的曖昧性(名詞句(NP)と指示形容詞(DA))により、図1に示す2つの構文構造(a),(b)が作成される。ここで(a)が選択されたとすると「彼女に」が生成されるが、次の語“aunt”の入力により、その選択は誤りであることが判明する。このため、(b)を選択し直し変換処理を実行すれば、「彼女に」を「彼女の叔母に」と言い直すことができる。

漸進的な翻訳において言い直しを活用することは有効であるが、一方で、あまりに頻繁に言い直しを生成すると翻訳結果の品質が低下し、意味の通らない翻訳文が生成されることになる。特に、入力文が長くなると、(1)構文的曖昧性が大きくなり、作成される構文構造の数が増大するため、構造を誤って選択する可能性が増す、さらに、(2)構文構造の選択機会が増えるため、選択を誤る機会も増加する、といった理由により言い直しの生成頻度が増える。実際に、英語会話文278文を用いた翻訳正解率に関する実験では、翻訳誤りと判定された50文の66%に相当する33文は、言い直しの過度の生成が誤りの原因であった [1]。

翻訳精度の向上のために言い直しの生成頻度を減らすことは重要であるが、そのためには変換処理における構文構造の選択誤りを削減する必要がある。その方法として、語単位での解析・変換処理という制約を、漸進性を損なわない程度に緩めることが考えられる。本稿では、(1)に対しては、解析部において入力語の先読みを許すこと、また(2)に対しては、変換部における構文構造の選択機会を数語おきにする、を検討する。

3 入力語の先読みによる言い直しの削減

漸進的な構文解析手法 [3] は、語が入力されるたびにそれまでの入力に対する可能な構文構造を作成する枠組である。そこで、次の入力語を先読みすれば、作成された構文構造の適否を文法上の制約から判定することができ、不適切な構造を削除し、変換処理対象となる選択肢を減らすことができる。例えば、文(2.1)の“her”に対して処理を実行する時点で、次の語“aunt”を先読みすれば、図1の(a)を削除することができる。よって、(b)が選択され「彼女の」が出力されることになり、言い直しは生じない。1語先読みする解析手法に基づ



(a)herを名詞句とする構造 (b)herを指示形容詞とする構造

図1: “her”の入力時点で作成される構文

表1: 先読みを用いた翻訳処理の正解率

	平均言い直し回数	翻訳正解率(%)
先読み無し	1.06	82.0
1語先読み	0.58	88.5

き、文献 [1] の実験と同様の英語対話文を用いて、言い直し回数と翻訳正解率を表1に示す。先読みを行うことで、言い直し回数が減り、翻訳正解率が向上することを確認した。

4 構文構造の選択機会の減少に

構文構造の選択を数語おきに行えば、言い直しの発生を抑制できる。これは、翻訳処理単位を語から数語単位にすることで、より現実的な方法であるといえる。本稿では、この手法を実行すると、そのタイミングは語“her”が入力された時点となる。よって、構文構造の選択機会が減る。語“aunt”に対しては「彼女の叔母に」(1語、2語、3語、及び4語)おき、すなわち1語おきの変換処理方式に基づき、前節と同様に構文構造の選択機会と翻訳正解率を図2に示す。処理単位が数語おきになることを確認した。

5 まとめ

漸進的な話し言葉翻訳において、処理単位の拡大を検討した。本稿で用いた2つの手法は、それぞれ異なる観点から翻訳結果作成の即時性の度合はやや低下するものの、翻訳精度の向上を達成することを確認した。

参考文献

- [1] 松原, 浅井, 外山, 稲垣: 不適格表現の翻訳処理に関する検討, 言語学論誌 C, Vol.118, No.1, pp.71-77.
- [2] Matsubara, S. and Inagaki, Y.: Incremental Chart-based Translation, *IEICE Trans. Inf. & Comm.*
- [3] Matsubara, S. et al.: Chart-based Incremental Chart-based Language Translation, *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing*